Creating PDF Documents
August 21, 2001

The creation of Portable Document Format (PDF) documents is essential for the electronic filing component of CM/ECF; it is the only format that the application accepts. The CM/ECF project team has evaluated various options for creating PDF documents; findings and guidelines are presented below.

Background

There are two primary methods for creating PDF documents: formatting text documents into PDF at the time of creation or scanning imaged documents from paper into PDF. The former method offers at least four significant advantages over the latter.

> *Reduced Labor*: Imaging is a labor-intensive task and if documents must be imaged, either the attorney or the court will have to provide that labor. Courts should strive to limit the amount of imaging done by court staff.
>
> *Improved System Performance*: Text documents are much smaller than imaged documents. For a typical document, its PDF text version would be only 20% of the size of its imaged version. Therefore, whenever an imaged document is stored or viewed, it puts five times more load on the network than it would as a text document. It also take five times as long to transfer. That extra transfer time is particularly noticeable when the user is working over the DCN or the Internet.
>
> *Reduced System Storage Cost*: An imaged document consumes about five times more storage space than it would as a text file.
>
> *Text Search Capability*: Text documents can be searched for words and phrases (as is done in Westlaw or Lexis). Imaged documents cannot be searched without first performing an extra step called optical character recognition (OCR). OCR is labor-intensive and does not yield 100% accuracy.

Creating a fully electronic case file is a worthy goal, and to do that some imaging is necessary. However, it might not be possible to convert some exhibits and other documents into an electronic format. For others, it might be possible but not economic. Courts need to develop their own policy regarding the handling of non-electronic documents, and in doing so, consider the cost versus the benefit of having court staff scan particular types of documents. For example, if a document is not likely to be viewed, or will be very seldom viewed, there may be little benefit to offset the cost of scanning it.

Creating PDF documents from a word processing package

The optimal method for creating a PDF document for filing in CM/ECF is a simple method: create it directly from a word processing application using Adobe Acrobat's PDFWriter. PDF files created in this way have two advantages: they are much smaller in size than documents that have been scanned and they are text-searchable by court users. The CM/ECF project team recommends this method for creating a PDF file from a word processing application. (Similar commercial PDF products such as "activePDF" are under review, but are not yet recommended.)

Adobe Acrobat

The price of Adobe Acrobat is approximately $250; the attorney discount has been discontinued. However, for law firms and courts that need to purchase more than 10 copies, Adobe provides a volume discount; see the http://www.adobe.com/store/openoptions/main.html web site for more information. Each court is given $500 for the purchase of Adobe Acrobat; the procurement of a judiciary-wide enterprise license is under consideration. Although the price of this package has risen somewhat, the CM/ECF team continues to recommend it as the best and most efficient way to create a PDF file.

PDFWriter

PDFWriter is part of the Adobe Acrobat package. The CM/ECF project team recommends this tool for use in creating a PDF document from a word processing application because it:

- creates a file that is smaller in size than a scanned document
- creates a file that is text searchable
- converts the document more quickly than Distiller (see below)

To ensure that the formatting and appearance of the document remain the same when viewed through the word processor and when viewed or printed through the PDF reader, the printer (File/Print menu) must be set to "Acrobat PDFWriter" *before* beginning to compose or edit the document. If a document is initially prepared with some other printer specified, the ultimate conversion to PDF is very likely to introduce changes in pagination, fonts, spacing, or other formatting elements, requiring further proofreading and further editing. Once the document is saved as a PDF file, always print from the PDF reader (rather than from the word processor) to be sure that the printed copy matches the court's official copy.

If Acrobat PDFWriter is set as the default printer, no other steps should be needed. If some other printer is set as the default, Acrobat PDFWriter must be selected as the current printer immediately after opening the word processing application (or immediately after choosing to

create a new document). In WordPerfect, each time the document is opened for editing before the final version is ready for filing, the printer should be set again to Acrobat PDFWriter. In Word, Acrobat PDFWriter is retained as the printer until the Word application is closed; if just the document has been closed and then reopened, the printer will still be set to Acrobat PDFWriter.

The latest version, 5.0, has been tested by the CM/ECF project team; there is no difference between this and Adobe versions 3.0 and 4.0 in either docketing the PDF documents into the application or in accessing the documents from the application. Also, documents created in the 5.0 version can be accessed by earlier versions of Adobe Reader. The CM/ECF project team also tested the integrity of 32 fonts when converted from either WordPerfect version 8 or Word 97 into PDF using Adobe Acrobat 5.0 PDFWriter. The tests showed that while characteristics of some fonts were lost in the conversion, every font conversion produced legible results. Courts may want to notify users about these fonts. The details of the tests appear in Appendix B to this document.

Distiller

This tool can also be used to create a PDF document but it is not recommended unless there is a need for embedded images in the document because there are a couple of drawbacks to using this tool:

- it creates a file significantly larger than that created by PDFWriter
- it takes significantly longer for the file to be created

Note that the default installation of Adobe Acrobat 5.0 includes the installation of Distiller but not PDFWriter. To load PDFWriter, you must choose the "custom" installation. (This procedure is different from the installations of Adobe 3.0 and 4.0, which included PDFWriter in the default installation.)

WordPerfect

Although WordPerfect versions 9 and 10 offer a feature ("Publish-to-PDF") that allows a document to be converted directly into PDF format, it is not recommended; the file created in this way is unnecessarily large.

Testing by the CM/ECF project team shows that an eight page document converted to PDF with WordPerfect 9.0's Publish-to-PDF feature is three times larger in size than the same document converted with Acrobat PDFWriter. When a one page document with an image embedded in it (a small agency seal) is converted, the difference in size is much greater; the PDF file created by WordPerfect is nearly 100 times larger than the file created with

PDFWriter. This difference in size can be lessened by changing the bitmap compression setting from the default of ZIP to JPEG (through the Details tab) and the quality factor from 2 (the highest quality) to 255 (the lowest quality). However, even with the quality set to 255, the file is still three times greater and with the quality set to 2, the file is 11 times greater. These tests were done with the Advanced Output setting set to the default of RGB; changing this setting to gray does not significantly change the size of the file. The quality of the seal is better at the highest quality setting (and better than the quality of the seal created through PDFWriter).

Preliminary testing of this feature in WordPerfect 10 shows that improvements have been made in the size of the file created. However, compared to the file created by PDFWriter, the WordPerfect-created PDF file is still nearly three times greater for a plain text file and eight times greater for a file with an image embedded in it. The size of the imaged file can be diminished to about four times greater by changing settings.

Microsoft Word

Microsoft Word 97, Word 2000, and Word 2002 do not currently provide a PDF converter and the Product Guide for Microsoft Office XP makes no reference to it. There is, however, a macro installed with Adobe Acrobat (beginning with the 3.01 version) called PDFMaker that provides enhanced features for creating PDF files from Word. It is installed by Acrobat in the Microsoft Office/Office/Startup folder and is accessed through the File/Create Adobe PDF menu item or through an Adobe icon on the toolbar. It uses Acrobat PDFWriter or Distiller and converts Word features such as headings to PDF bookmarks, URLs to PDF Weblinks, cross-references within a document to PDF links, etc. A full list of these features can be found at the www.adobe.com web site (search for PDFMaker) or for users that have already installed Adobe Acrobat, a help file can be found in the Program Files/Adobe/Acrobat*x*.0/Help/ENU folder. (Adobe Acrobat 4.0 users who wish to use PDFMaker with Word 2000 must upgrade [this is a free upgrade from Adobe] to Acrobat 4.05; the PDFMaker that shipped with Acrobat 4.0 does not work with Word 2000.) PDFMaker is not available for the Macintosh operating system.

The PDF files created in Word using the File/Print to Acrobat PDFWriter method and the File/Create Adobe PDF (the PDFMaker macro) method are nearly identical in size. There do not appear to be any differences in the quality of the PDF files produced, but it has not been fully tested by the CM/ECF project team. The advantage of the PDFMaker macro is the additional conversion features that it provides for Word documents. If these particular Word features are not used, the File/Print to Acrobat PDFWriter method is adequate.

Creating PDF documents from a scanning system

For those documents that must be imaged because a word-processed version does not exist, the preferred method is to scan the document directly into PDF format using Adobe Acrobat (both 4.0 and 5.0 provide this feature). (From the File menu, choose Import/Scan.) Scanning to a format other than PDF (e.g., TIFF) would add both delay and labor, requiring not only the scan but then a conversion to PDF.

Factors to consider when scanning

There are several factors to consider when determining how the scanning will be done.

The quality of the document when scanned is determined by the level of detail recorded by the scanner. This detail is referred to as the resolution, which is measured by the dots per inch (dpi). A higher resolution:

- is slower to scan (this is dependent also on the scanner and the number of pages that are scanned).
- creates a document with a larger file size. This, in turn, causes the loading of the document into the case management application and the retrieval of the document from the application, to be slower.

There is a trade-off between the resolution, the speed at which the document is scanned, and the file size of the scanned document. For example, using a Ricoh IS-430 scanner (rated at 30-45 pages per minute), a 100 page document scanned at 300 dpi produces a file of 6.97 MB and takes 3 minutes and 51 seconds to scan. The same document, scanned at 150 dpi produces a file of 3.67 MB and takes 2 minutes, 15 seconds to scan.

It is therefore important to find a resolution that will provide a high quality document, with a file size that does not hamper the length of time it takes to scan, load, and retrieve the document. Of those surveyed, most courts that scan recommend a resolution between 200-300 dpi. The CM/ECF project team recommends 200 dpi resolution.

Another factor to consider is the mode of scanning. Almost always, it should be done in black and white, so that the file size will be as small as possible. However, there are some documents (those with shaded boxes, for example) that may need grayscale instead; this, however, will produce a file that is much larger in size than the document scanned in black and white. Never scan in color, unless absolutely necessary (perhaps for exhibits originally produced in color). This, too, will produce a document with an extremely large file size, which will take longer to load into the application and longer to retrieve for viewing purposes.

Limits on size

Because of the staff time, disk space, and long retrieval time consumed by large documents, some courts (including non-CM/ECF courts) have set limits on either the number of pages that will be accepted electronically or the file size that will be accepted. These limits range from 25 to 150 pages, and from 1.5 to 3 MB; if a document exceeds the limit, the policy is either to break the document into separate, smaller documents or to file and maintain the document in paper form with the court. Often when this latter method is chosen, a one page PDF document is filed as a "placeholder" for the larger document, directing the reader to the location of the whole document (e.g., the reader must come to the courthouse to see the document). While this method leads to a mix of paper and electronic files in some cases, the practical costs of scanning voluminous documents makes a pure electronic case file impractical in these cases. The CM/ECF project team has contacted the judiciary's records management staff regarding the long-term implications of mixed paper and electronic cases for archiving and storage purposes. Their guidance is expected soon and will be provided to the CM/ECF courts when it is published.

The CM/ECF project team is in the process of developing a new program that will allow each court to determine the maximum size for a PDF file that is loaded into the court's application; the project team recommends 1.5 or 2 MB. When a user attempts to load a file larger than this, the court will have the option of either allowing the user to continue or forcing the user to stop. (The court will also create its own message informing the user of the action that (s)he must take.) This new program will be distributed to all current CM/ECF courts in an Emergency Modification Request (EMR) release when it is completed and will be included in the application software for new courts.

Another method for handling large scanned documents is to scan only the relevant portions. Model local rules for electronic case filing regarding the issue of scanning large documents for both district and bankruptcy courts have been developed by the Judicial Conference Committee on Court Administration and Case Management's (CACM) Subcommittee on Electronic Filing Rules. The subcommittee's proposal, which has been approved by CACM, but not yet by the Judicial Conference, addresses the issue in its Rule 5, Attachments and Exhibits. Adapted from the Southern District of New York bankruptcy procedures, it states in part, that "A Filing User must submit as exhibits or attachments only those excerpts of the referenced documents that are directly germane to the matter under consideration by the court." Commentary number 2 on this rule notes that

> It is often the case that only a small portion of a much larger document is relevant to the matter before the court. In such cases, scanning the entire document imposes an inappropriate burden on both the litigants and the courts. To alleviate some of this inconvenience, the Model Rule provides

that a Filing User must submit as the exhibit only the relevant excerpts of a larger document. The responding party then has a right to submit other excerpts of the same document under the principle of completeness.

OCR, Adobe Capture, and Kofax Ascent Capture

A document that has been scanned using an OCR feature will be text-searchable, unlike the pure image file. However, the drawback to creating a document through the OCR process is that it is generally only about 95% - 97% accurate. Different packages claim different accuracy rates, but no package claims 100% accuracy. Even a 3% error rate can add up to a fair number of errors, depending on the size of the document, and fixing each error can substantially increase the amount of time needed to prepare the documents for loading into the application.

Adobe has a package called Capture that allows conversion of scanned pages to PDF files that are text searchable, using an OCR process. However, as with all documents produced through OCR, correcting the errors slows the process. It does, however, create a PDF file that is substantially smaller in size than the scanned document (estimated at 1/16 the size of the scanned file).

The Washington Western bankruptcy court has developed a batch scanning system using Kofax Ascent Capture, software that scans a document into TIFF format (this package does not scan directly into PDF), automatically converts the TIFF file to PDF (or TXT if the file is a creditor matrix), and stores the file on a network (one license of this software is required for each scanner on which it is used). The court also purchased professional services (including the source code) to create a scanning turnkey solution (which includes barcode technology [the barcode is used to separate the batches] and creditor matrix viewer). For more information about this system, contact Tim AhYat at 206-553-7545.

Converting scanned documents from existing imaging system to PDF

For the ICMS and NIBS courts that are currently using an imaging system that creates documents in TIFF, there is a freeware program available called *c42pdf.exe* that will convert these files into PDF. (After the conversion, the new PDF file is virtually the same size as the original TIFF file.) This program has been successfully used by the California Southern bankruptcy court to convert its imaged documents (contact Russ Reynolds at 619-557-6213 for more information about how it is used). The program can be downloaded from the www.pdfzone.com or www.activepdf.com web sites, along with the source code and instructions. The converted files can then be loaded into the CM/ECF application when the data in the legacy applications is converted to CM/ECF.

Adobe Capture can also be used to convert TIFF files to PDF files.

Webinator

In order to be able to search for particular words in the PDF documents that were created to be text-searchable, CM/ECF uses a program called Webinator, which is installed on the local server. It is typically placed on the Utilities menu and available only to court users. (It is available only to court users because it runs directly on the inside/DCN-accessible server, to which the attorneys do not have access.)

Border Manager (and other proxy servers)

Border Manager is a Novell package that allows courts to manage user access to both an Intranet and the Internet. While investigating a reported problem with performance that was thought initially to be related to imaging practices, it was discovered that a proxy server had dramatically slowed the system performance. The problem was eliminated by creating a script to route the local CM/ECF traffic so that it bypasses the proxy server. (There is no benefit in routing CM/ECF traffic through the proxy server.) This script is enabled in Netscape through the "Automatic proxy configuration" setting in Preferences/Advanced/Proxies or in Internet Explorer, through the "Automatic configuration script" setting. It can be found on the CM/ECF Application Product Pages maintained by the Systems Delivery and Support Division in San Antonio.

Right click on the push pin to extract the script.

Appendix A

The following is a list of scanners that are currently being used by CM/ECF courts.

Scanner model and speed	Court
Clerk's office: Fujitsu 3097, 40 pages/minute (PPM) ($4635); Fujitsu 4097 ($4,728) *Chambers*: HP ScanJet 6350, a few pages/minute ($499)	OH-N District
Clerk's office: Fujitsu 3097, 40 PPM ($4635); Fujitsu 4097 ($4,728) *Chambers*: HP ScanJet 6350, a few pages/minute ($499)	DC District
Fujitsu M4097D, 50 PPM simplex and 90 PPM duplex	MI-W District
Large format scanner: Bell&Howell 8125D, 125 PPM duplex *Medium format scanner*: Kodak 1500/2500, 60 PPM *Smaller scanners*: Visioneer Strobe Pro, 2 PPM	NY-E District
HP Network ScanJet5, 15-20 PPM	OR District
Ricoh, model IS-430 with SCSI interface and ADF, 30-45 PPM	VA-E Bankruptcy
Fujitsu 93GX, 25 PPM	CA-S Bankruptcy
Desktop: 3091D with 30 page ADF, 15 PPM (government price is approximately $700) *High capacity scanners*: Ricoh IS420, 33 PPM and Ricoh IS450D, 66 PPM; both with 100 page ADF	VT Bankruptcy
Fujitsu 4097D, M4097G, M3097DG, 25-50 PPM	NC-W Bankruptcy
Canon DR-5020	DE Bankruptcy
Large scanning jobs: three Ricoh IS450DE, with multi-page feeders, 57 PPM *Intake clerks and case administrators*: Visioneer Paperport Strobe Pro, single-sheet feeder, 15 PPM	AZ Bankruptcy
Ricoh 420, 38 PPM and Ricoh 450DE, 57 PPM	GA-N Bankruptcy
Panasonic KV-SS55EX, 40 PPM; Panasonic S2055, 50 PPM; flatbed HP Scanjets for miscellaneous documents	NY-S Bankruptcy

Visioneer One Touch 8650 desktop, 4 PPM HP Scanjet V Si group network, 15 PPM HP 9100C Digital Sender network, 15 PPM Fujitsu 3750 workstation, 37 PPM (requires network PC) Savin 9945 multi-use copier network, 80 PPM	TX-W Bankruptcy
Canon DR5020, 80-90 PPM, 500 page paper tray ($5600)	WA-W Bankruptcy

Appendix B

Adobe Acrobat 5.0 was tested for its ability to maintain font integrity through conversions to PDF of documents created in WordPerfect version 8 and Microsoft Word 97. All of the 32 fonts tested were done with a 12 point size. The results show that while characteristics of some fonts were lost in the conversion (e.g., appeared smaller), every font conversion produced legible results and none produced a problem with either character spacing or character overlap (superimposition).

Fonts that maintained integrity in the conversion from WordPerfect 8 to PDF

Albertus Extra Bold	Courier New
Albertus Medium Bold	Haettenschweiler
Antique Olive	Impact
Arial	Letter Gothic
Arial Black	Letter Gothic MT
Arial NarrowBook Antiqua	Line Printer
Bookman Old Style	Marigold
Century Gothic	Modern
Century Schoolbook	Symbol (2 styles)
CG Omega	Tahoma
CG Times	Times New Roman
Clarendon Condensed Bold	Times New Roman (bold)
Comic Sans MS	Univers
Coronet	Univers Condensed Regular
Courier	Verdana

Fonts that were altered in the conversion from WordPerfect 8 to PDF

Coronet	appears much larger and in a different font, an Adobe default font
Line Printer	appears much larger and in a different font, an Adobe default font
Marigold	appears much larger and in a different font, an Adobe default font

Fonts that maintained integrity in the conversion from Word 97 to PDF

Arial	Garamond
Arial Black	Haettenschweiler
Arial NarrowBook Antiqua	Impact
Bookman Old Style	Letter Gothic
Century Gothic	Symbol (2 styles)
Century Schoolbook	Tahoma
CG Times	Times New Roman
Comic Sans MS	Univers
Courier	Univers Condensed Regular
Courier New	Verdana

Fonts that were altered in the conversion from Word 97 to PDF

Albertus Extra Bold	bold characteristic lost, appears in a different font, an Adobe default font
Albertus Medium Bold	appears in a different font, an Adobe default font
Antique Olive	appears much smaller and in a different font, an Adobe default font
CG Omega	appears much smaller and in a different font, an Adobe default font
Clarendon Condensed Bold	appears much smaller and in a different font, an Adobe default font
Coronet	appears much larger and in a different font, an Adobe default font
Line Printer	appears much larger and in a different font, an Adobe default font
Marigold	appears much larger and in a different font, an Adobe default font
Modern	appears in a different font, an Adobe default font
Times New Roman (bold)	appears in a different font, an Adobe default font